

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 09/719,533E
Source: IFW16
Date Processed by STIC: 10/27/05

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:10

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

SEQUENCE LISTING

```

3 (1) GENERAL INFORMATION:
5     (i) APPLICANT: Oon, Chong Jin
6             Lim, Gek Keow
7             Leong, Ai Lin
8             Zhao, Yi
9             Chen, Wei Ning
11    (ii) TITLE OF INVENTION: A VACCINE-INDUCED HEPATITIS B VIRAL
12             STRAIN AND USES THEREOF
14    (iii) NUMBER OF SEQUENCES: 11
16    (iv) CORRESPONDENCE ADDRESS:
17            (A) ADDRESSEE: Ladas & Parry
18            (B) STREET: 26 West 61 Street
19            (C) CITY: New York
20            (D) STATE: New York
21            (E) COUNTRY: USA
22            (F) ZIP: 10023
24    (v) COMPUTER READABLE FORM:
25            (A) MEDIUM TYPE: Floppy disk
26            (B) COMPUTER: IBM PC compatible
27            (C) OPERATING SYSTEM: PC-DOS/MS-DOS
28            (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
30    (vi) CURRENT APPLICATION DATA:
C--> 31            (A) APPLICATION NUMBER: US/09/719,533E
C--> 32            (B) FILING DATE: 10-Jul-2001
33            (C) CLASSIFICATION: 435
35    (vii) PRIOR APPLICATION DATA:
36            (A) APPLICATION NUMBER: PCT/SG98/00045
37            (B) FILING DATE: 19-JUN-1998
39    (viii) ATTORNEY/AGENT INFORMATION:
40            (A) NAME: Mass, Clifford J.
41            (B) REGISTRATION NUMBER: 30,086
42            (C) REFERENCE/DOCKET NUMBER: U-013108-9
44    (ix) TELECOMMUNICATION INFORMATION:
45            (A) TELEPHONE: (212) 708-1890
47 (2) INFORMATION FOR SEQ ID NO: 1:
49     (i) SEQUENCE CHARACTERISTICS:
50            (A) LENGTH: 3215 base pairs
51            (B) TYPE: nucleic acid
52            (C) STRANDEDNESS: double
53            (D) TOPOLOGY: circular
55     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
57 CTCCACCACT TTCCACCAAA CTCTTCAAGA TCCCAGAGTC AGGGCCCTGT ACTTTCCTGC

```

60

RAW SEQUENCE LISTING

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:10

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

59	TGGTGGCTCC	AGTTCAGGAA	CAGTGAGCCC	TGCTCAGAAT	ACTGTCTCTG	CCATATCGTC	120
61	AATCTTATCG	AAGACTGGGG	ACCCTGTACC	GAACATGGAG	AACATCGCAT	CAGGACTCCT	180
63	AGGACCCCTG	CTCGTGTTAC	AGGCGGGGTT	TTTCTTGTTG	ACAAAAATCC	TCACAATACC	240
65	GCAGAGTCTA	GACTCGTGGT	GGACTTCTCT	CAATTTTCTA	GGGGGAACAC	CCGTGTGTCT	300
67	TGGCCAAAAT	TCGCAGTCCC	AAATCTCCAG	TCACTCACCA	ACCTGTTGTC	CTCCAATTTG	360
69	TCCTGGTTAT	CGCTGGATGT	GTCTGCGGCG	TTTTATCATC	TTCTCTGCA	TCCTGCTGCT	420
71	ATGCCTCATC	TTCTTGTTGG	TTCTTCTGGA	CTATCAAGGT	ATGTTGCCCG	TTTGTCTCT	480
73	AATTCCAGGA	TCAACAACAA	CCAGCACCGG	ACCATGCAAA	ACCTGCACAA	CTCCTGCTCA	540
75	AGGAACCTCT	ATGTTTCCCT	CATGTTGCTG	TACAAAACCT	ACGGACAGAA	ACTGCACCTG	600
77	TATTCCCATC	CCATCATCTT	GGGCTTTCGC	AAAATACCTA	TGGGAGTGGG	CCTCAGTCCG	660
79	TTTCTCTTGG	CTCAGTTTAC	TAGTGCCATT	TGTTCAAGTG	TTCTAGGGG	TTTCCCCCAC	720
81	TGTCTGGCTT	TCAGTTATAT	GGATGATGTG	GTTTTGGGGG	CCAAGTCTGT	ACAACATCTT	780
83	GAGTCCCTTT	ATGCCGCTGT	TACCAATTTT	CTTTTGTCTT	TGGGTATACA	TTTAAACCTT	840
85	CACAAAACAA	AAAGATGGGG	ATATTCCCTT	AACCTTCATG	GATATGTCAT	TGGGAGTTGG	900
87	GGCACATTGC	CACAGGAACA	TATTGTACAA	AAAATCAAAA	TGTGTTTTAG	GAAACTTCCT	960
89	GTAAACAGGC	CTATTGATTG	GAAAGTATGT	CAACGAATTG	TGGGTCTTTT	GGGGTTTGCC	1020
91	GCCCCTTTCA	CGCAATGTGG	ATATCCTGCT	TTAATGCCTT	TATATGCATG	TATACAAGCA	1080
93	AAACAGGCTT	TTACTTTCTC	GCAAACCTTAC	AAGACCTTTC	TAAGTAAACA	GTATCTGAAC	1140
95	CTTTACCCCG	TTGCTCGGCA	ACGCCCTGGT	CTGTGCCAAG	TGTTTGCTGA	CGCAACCCCC	1200
97	ACTGGTTGGG	GCTTGGCCAT	AGGCCATCAG	CGCATGCGTG	GAACCTTTGT	GTCTCCTCTG	1260
99	CCGATCCATA	CTGCGGAACT	CCTAGCCGCT	TGTTTTGCTC	GCAGCAGGTC	TGGGGCAAAA	1320
101	CTCATCGGGA	CTGACAATTC	TGTCGTGCTC	TCCCGCAAGT	ATACATCATT	TCCATGGCTG	1380
103	CTAGGCTGTG	CTGCCAACTG	GATCCTGCGC	GGGACGTCTT	TTGTTTACGT	CCCGTCGGCG	1440
105	CTGAATCCCG	CGGACGACCC	CTCCCGGGGC	CGCTTGGGGC	TCTACCGCCC	GCTTCTCCGC	1500
107	CTGTTATACC	GACCGACCAC	GGGGCGCACC	TCTCTTTACG	CGGACTCCCC	GTCTGTGCTT	1560
109	TCTCATCTGC	CGGACCGTGT	GCACTTCGCT	TCACCTCTGC	ACGTCGCATG	GAGACCACCG	1620
111	TGAACGCCCA	CGGGAACCTG	CCCAAGGTCT	TGCATAAGAG	GACTCTTGGA	CTTTCAGCAA	1680
113	TGTCAACGAC	CGACCTTGAG	GCATACTTCA	AAGACTGTGT	GTTTAATGAG	TGGGAGGAGT	1740
115	TGGGGGAGGA	GGTTAGGTTA	AAGGTCTTTG	TACTAGGAGG	CTGTAGGCAT	AAATTGGTGT	1800
117	GTTCAACCATC	ACCATGCAAC	TTTTTTCACCT	CTGCCTAATC	ATCTCATGTT	CATGTCCTAC	1860
119	TGTTCAAGCC	TCCAAGCTGT	GCCTTGGGTG	GCTTTGGGGC	ATGGACATTG	ACCCGTATAA	1920
121	AGAATTTGGA	GCTTCTGTGG	AGTTACTCTC	TTTTTTGCCT	TCTGACTTTT	TTCTTCTAT	1980
123	TCGAGATCTC	CTCGACACCG	CCTCTGCTCT	GTATCGGGAG	GCCTTAGAGT	CTCCGGAACA	2040
125	TTGTTACCTT	CACCATACGG	CACTCAGGCA	AGCTATTCTG	AGTTGGGGTG	AGTTAATGAA	2100
127	TCTAGCCACC	TGGGTGGGAA	GTAATTTGGA	AGATCCAGCA	TCCAGGGAAT	TAGTAGTCAG	2160
129	CTATGTCAAC	GTTAATATGG	GCCTAAAAAT	CAGACAACTA	TTGTGGTTTC	ACATTTCTCTG	2220
131	TCTTACTTTT	GGGAGAGAAA	CTGTTCTTGA	ATATTTGGTG	TCTTTTGGAG	TGTGGATTCTG	2280
133	CACTCCTCCT	GCATATAGAC	CACCAAATGC	CCCTATCTTA	TCAACACTTC	CGGAAACTAC	2340
135	TGTTGTTAGA	CGAAGAGGCA	GGTCCCCTAG	AAGAAGAACT	CCCTCGCCTC	GCAGACGAAG	2400
137	GTCTCAATCG	CCGCGTCGCA	GAAGATCTCA	ATCTCGGGAA	TCTCAATGTT	AGTATTCCTT	2460
139	GGACACATAA	GGTGGGAAAC	TTTACGGGGC	TTTATTCTTC	TACGGTACCT	TGCTTTAATC	2520
141	CTAAATGGCA	AACTCCTTCT	TTTCCGGACA	TTCATTTGCA	GGAGGACATT	CTTGATAGAT	2580
143	GTAAGCAATT	TGTGGGGCCC	CTTACAGTAA	ATGAAAACAG	GAGACTAAAA	TTAATTATGC	2640
145	CTGCTAGGTT	TTATCCAAAT	GTTACTAAAT	ATTTGCCCTT	AGATAAAGGG	ATCAAACCAT	2700
147	ATTATCCAGA	GTATGTAGTT	AATCATTACT	TCCAGACGCG	ACATTATTTA	CACACTCTTT	2760
149	GGAAGGCGGG	GATCTTATAT	AAAAGAGAGT	CCACACGTAG	CGCCTCATTT	TGCGGGTCAC	2820
151	CATATTCTTG	GGAACAAGAT	CTACAGCATG	GGAGGTTGGT	CTTCCAAACC	TCGAAAAGGC	2880
153	ATGGGGACAA	ATCTTTCTGT	CCCCAATCCC	CTGGGATTCT	TCCCCGATCA	TCAGTTGGAC	2940
155	CCTGCATTCA	AAGCCAACTC	AGAAAATCCA	GATTGGGACC	TCAACCCGCA	CAAGGACAAC	3000

RAW SEQUENCE LISTING

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:10

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

```

157 TGGCCGGACG CCAACAAGGT GGGAGTGGGA GCATTCGGGC CAGGGTTCAC CCCTCCTCAT 3060
159 GGGGGACTGT TGGGGTGGAG CCCTCAGGCT CAGGGCCTAC TCACAACTGT GCCAGCAGCT 3120
161 CCTCCTCTG CCTCCACCAA TCGGCAGTCA GGAAGGCAGC CTACTCCCTT ATCTCCACCT 3180
163 CTAAGGGACA CTCATCCTCA GGCCATGCAG TGGAA 3215

```

166 (2) INFORMATION FOR SEQ ID NO: 2:

168 (i) SEQUENCE CHARACTERISTICS:

169 (A) LENGTH: 843 amino acids

170 (B) TYPE: amino acid

171 (C) STRANDEDNESS: single

172 (D) TOPOLOGY: linear

176 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

178 Met Pro Leu Ser Tyr Gln His Phe Arg Lys Leu Leu Leu Leu Asp Glu
179 1 5 10 15
181 Glu Ala Gly Pro Leu Glu Glu Glu Leu Pro Arg Leu Ala Asp Glu Gly
182 20 25 30
184 Leu Asn Arg Arg Val Ala Glu Asp Leu Asn Leu Gly Asn Leu Asn Val
185 35 40 45
187 Ser Ile Pro Trp Thr His Lys Val Gly Asn Phe Thr Gly Leu Tyr Ser
188 50 55 60
190 Ser Thr Val Pro Cys Phe Asn Pro Lys Trp Gln Thr Pro Ser Phe Pro
191 65 70 75 80
193 Asp Ile His Leu Gln Glu Asp Ile Leu Asp Arg Cys Lys Gln Phe Val
194 85 90 95
196 Glu Pro Leu Thr Val Asn Glu Asn Arg Arg Leu Lys Leu Ile Met Pro
197 100 105 110
199 Ala Arg Phe Tyr Pro Asn Val Thr Lys Tyr Leu Pro Leu Asp Lys Gly
200 115 120 125
202 Ile Lys Pro Tyr Tyr Pro Glu Tyr Val Val Asn His Tyr Phe Gln Thr
203 130 135 140
205 Arg His Tyr Leu His Thr Leu Trp Lys Ala Gly Ile Leu Tyr Lys Arg
206 145 150 155 160
208 Glu Ser Thr Arg Ser Ala Ser Phe Cys Gly Ser Pro Tyr Ser Trp Glu
209 165 170 175
211 Gln Asp Leu Gln His Gly Arg Leu Val Phe Gln Thr Ser Lys Arg His
212 180 185 190
214 Gly Asp Lys Ser Phe Cys Pro Glu Ser Pro Gly Ile Leu Pro Arg Ser
215 195 200 205
217 Ser Val Gly Pro Cys Ile Gln Ser Gln Leu Arg Lys Ser Arg Leu Gly
218 210 215 220
220 Pro Gln Pro Ala Gln Gly Gln Leu Ala Gly Arg Gln Gln Gly Gly Ser
221 225 230 235 240
223 Gly Ser Ile Arg Ala Arg Val His Pro Ser Ser Trp Gly Thr Val Gly
224 245 250 255
226 Val Glu Pro Ser Gly Ser Gly Pro Thr His Asn Cys Ala Ser Ser Ser
227 260 265 270
229 Ser Ser Cys Leu His Gln Ser Ala Val Arg Lys Ala Ala Tyr Ser Leu
230 275 280 285
232 Ile Ser Thr Ser Lys Gly His Ser Ser Ser Gly His Ala Val Glu Leu
233 290 295 300

```

RAW SEQUENCE LISTING

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:10

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

```

235   His His Phe Pro Pro Asn Ser Ser Arg Ser Gln Ser Gln Gly Pro Val
236   305                               310                               315                               320
238   Leu Ser Cys Trp Trp Leu Gln Phe Arg Asn Ser Glu Pro Cys Ser Glu
239                               325                               330                               335
241   Tyr Cys Leu Cys His Ile Val Asn Leu Ile Glu Asp Trp Gly Pro Cys
242                               340                               345                               350
244   Thr Glu His Gly Glu His Arg Ile Arg Thr Pro Arg Thr Pro Ala Arg
245                               355                               360                               365
247   Val Thr Gly Gly Val Phe Leu Val Asp Lys Asn Pro His Asn Thr Ala
248                               370                               375                               380
250   Glu Ser Arg Leu Val Val Asp Phe Ser Gln Phe Ser Arg Gly Asn Thr
251   385                               390                               395                               400
253   Arg Val Ser Trp Pro Lys Phe Ala Val Pro Asn Leu Gln Ser Leu Thr
254                               405                               410                               415
256   Asn Leu Leu Ser Ser Asn Leu Ser Trp Leu Ser Leu Asp Val Ser Ala
257                               420                               425                               430
259   Ala Phe Tyr His Leu Pro Leu His Pro Ala Ala Met Pro His Leu Leu
260                               435                               440                               445
262   Val Gly Ser Ser Gly Leu Ser Arg Tyr Val Ala Arg Leu Ser Ser Asn
263   450                               455                               460
265   Ser Arg Ile Asn Asn Asn Glu His Arg Thr Met Glu Asn Leu His Asn
266   465                               470                               475                               480
268   Ser Cys Ser Arg Asn Leu Tyr Val Ser Leu Met Leu Leu Tyr Lys Thr
269                               485                               490                               495
271   Tyr Gly Gln Lys Leu His Leu Tyr Ser His Pro Ile Ile Leu Gly Phe
272                               500                               505                               510
274   Arg Lys Ile Pro Met Gly Val Gly Leu Ser Pro Phe Leu Leu Ala Gln
275                               515                               520                               525
277   Phe Thr Ser Ala Ile Cys Ser Val Val Arg Arg Ala Phe Pro His Cys
278   530                               535                               540
280   Leu Ala Phe Ser Tyr Met Asp Asp Val Val Leu Gly Ala Lys Ser Val
281   545                               550                               555                               560
283   Gln His Leu Glu Ser Leu Tyr Ala Ala Val Thr Asn Phe Leu Leu Ser
284                               565                               570                               575
286   Leu Gly Ile His Leu Asn Pro His Lys Thr Lys Arg Trp Gly Tyr Ser
287                               580                               585                               590
289   Leu Asn Phe Met Gly Tyr Val Ile Gly Ser Trp Gly Thr Leu Pro Gln
290                               595                               600                               605
293   Glu His Ile Val Gln Lys Ile Lys Met Cys Phe Arg Lys Leu Pro Val
294   610                               615                               620
296   Asn Arg Pro Ile Asp Trp Lys Val Cys Gln Arg Ile Val Gly Leu Leu
297   625                               630                               635                               640
299   Gly Phe Ala Ala Pro Phe Thr Gln Cys Gly Tyr Pro Ala Leu Met Pro
300                               645                               650                               655
302   Leu Tyr Ala Cys Ile Gln Ala Lys Gln Ala Phe Thr Phe Ser Gln Thr
303                               660                               665                               670
305   Tyr Lys Thr Phe Leu Ser Lys Gln Tyr Leu Asn Leu Tyr Pro Val Ala
306   675                               680                               685
308   Arg Gln Arg Pro Gly Leu Cys Glu Val Phe Ala Asp Ala Thr Pro Thr

```

RAW SEQUENCE LISTING

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:10

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

```

309          690          695          700
311 Gly Trp Gly Leu Ala Ile Gly His Gln Arg Met Arg Gly Thr Phe Val
312 705          710          715          720
314 Ser Pro Leu Pro Ile His Thr Ala Glu Leu Leu Ala Ala Cys Phe Ala
315          725          730          735
317 Arg Ser Arg Ser Gly Ala Lys Leu Ile Gly Thr Asp Asn Ser Val Val
318          740          745          750
320 Leu Ser Arg Lys Tyr Thr Ser Phe Pro Trp Leu Leu Gly Cys Ala Ala
321          755          760          765
323 Asn Trp Ile Leu Arg Gly Thr Ser Phe Val Tyr Val Pro Ser Ala Leu
324          770          775          780
326 Asn Pro Ala Asp Asp Pro Ser Arg Gly Arg Leu Gly Leu Tyr Arg Pro
327 785          790          795          800
329 Leu Leu Arg Leu Leu Tyr Arg Pro Thr Thr Gly Arg Thr Ser Leu Tyr
330          805          810          815
332 Ala Asp Ser Pro Ser Val Pro Ser His Leu Pro Asp Arg Val His Phe
333          820          825          830
335 Ala Ser Pro Leu His Val Ala Trp Arg Pro Pro
336          835          840

```

340 (2) INFORMATION FOR SEQ ID NO: 3:

342 (i) SEQUENCE CHARACTERISTICS:

343 (A) LENGTH: 400 amino acids

344 (B) TYPE: amino acid

345 (C) STRANDEDNESS: single

346 (D) TOPOLOGY: linear

351 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

353 Met Gly Gly Trp Ser Ser Lys Pro Arg Lys Gly Met Gly Thr Asn Leu
354 1          5          10          15
356 Ser Val Pro Asn Pro Leu Gly Phe Phe Pro Asp His Gln Leu Asp Pro
357          20          25          30
359 Ala Phe Lys Ala Asn Ser Glu Asn Pro Asp Trp Asp Leu Asn Pro His
360          35          40          45
362 Lys Asp Asn Trp Pro Asp Ala Asn Lys Val Gly Val Gly Ala Phe Gly
363          50          55          60
365 Pro Gly Phe Thr Pro Pro His Gly Gly Leu Leu Gly Trp Ser Pro Gln
366          65          70          75          80
368 Ala Gln Gly Leu Leu Thr Thr Val Pro Ala Ala Pro Pro Pro Ala Ser
369          85          90          95
371 Thr Asn Arg Gln Ser Gly Arg Gln Pro Thr Pro Leu Ser Pro Pro Leu
372          100          105          110
374 Arg Asp Thr His Pro Gln Ala Met Gln Trp Asn Ser Thr Thr Phe His
375          115          120          125
377 Gln Thr Leu Gln Asp Pro Arg Val Arg Ala Leu Tyr Phe Pro Ala Gly
378          130          135          140
380 Gly Ser Ser Ser Gly Thr Val Ser Pro Ala Gln Asn Thr Val Ser Ala
381          145          150          155          160
383 Ile Ser Ser Ile Leu Ser Lys Thr Gly Asp Pro Val Pro Asn Met Glu
384          165          170          175
386 Asn Ile Ala Ser Gly Leu Leu Gly Pro Leu Leu Val Leu Gln Ala Gly

```

VERIFICATION SUMMARY

DATE: 10/27/2005

PATENT APPLICATION: US/09/719,533E

TIME: 12:09:11

Input Set : A:\u0131089 sequence.txt

Output Set: N:\CRF4\10272005\I719533E.raw

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]